

# Blind and fully constrained unmixing of hyperspectral images

Rita Ammanouil, André Ferrari, Cédric Richard, David Mary

Laboratoire Lagrange, Université de Nice Sophia-Antipolis, France

{rita.ammannouil, andre.ferrari, cedric.richard, david.mary}@unice.fr

## Abstract

This paper addresses the problem of blind and fully constrained unmixing of hyperspectral images. Unmixing is performed without the use of any dictionary, and assumes that the number of constituent materials in the scene and their spectral signatures are unknown. The estimated abundances satisfy the desired sum-to-one and nonnegativity constraints. Two models with increasing complexity are developed to achieve this challenging task, depending on how noise interacts with hyperspectral data. The first one leads to a convex optimization problem, and is solved with the Alternating Direction Method of Multipliers. The second one accounts for signal-dependent noise, and is addressed with a Reweighted Least Squares algorithm. Experiments on synthetic and real data demonstrate the effectiveness of our approach.

## I. INTRODUCTION

Hyperspectral imaging is a continuously growing area of remote sensing, which has received considerable attention in the last decade. Hyperspectral data provide spectral images over hundreds of narrow and adjacent bands, coupled with a high spectral resolution. These characteristics are suitable for detection and classification of surfaces and chemical elements in the observed images. Applications include land use analysis, pollution monitoring, wide-area reconnaissance, and field surveillance, to cite a few. When unmixing hyperspectral images [1], two types of pixels can be distinguished: the pure pixels and the mixed ones. Each pure pixel, also called endmember, contains the spectral signature of a constituent material in the scene, whereas a mixed pixel consists of a mixture of the endmembers. The fraction of each endmember in a mixed pixel is called abundance. Three consecutive tasks are usually required for unmixing: determining the number of endmembers, extracting the spectral signature of the endmembers, and estimating their abundances for every pixel in the scene. Several algorithms have been proposed to perform each stage separately. Virtual Dimensionality (VD) [2], followed by N-FINDR [3] and FCLS [4] is among the most widely used processing pipeline. Alternative methods jointly performs (part of) these tasks in order to solve the blind source separation problem [5]–[8].

In order to introduce our approach, we shall now describe the noise-free case first. Consider the linear mixing model where a mixed pixel is expressed as a linear combination of the endmembers weighted by their fractional abundances. In matrix form, we simply have:

$$\tilde{\mathbf{S}} = \mathbf{R}\mathbf{A} \quad (1)$$

where  $\tilde{\mathbf{S}} = (\tilde{\mathbf{s}}_1, \dots, \tilde{\mathbf{s}}_N)$ ,  $\mathbf{R} = (\mathbf{r}_1, \dots, \mathbf{r}_M)$ ,  $\mathbf{A} = (\mathbf{a}_1, \dots, \mathbf{a}_M)^\top$ ,  $\tilde{\mathbf{s}}_j$  is the  $L$ -dimensional spectrum of the  $j$ -th pixel,  $L$  is the number of frequency bands,  $\mathbf{r}_i$  is  $L$ -dimensional spectrum of the  $i$ -th endmember,  $M$  is the number of endmembers,  $\mathbf{a}_i$  is the  $N$ -dimensional abundance map of the  $i$ -th endmember, and  $N$  is the number of pixels in the image. Model (1) means that the  $(i, j)$ -th entry  $\mathbf{A}_{ij}$  of matrix  $\mathbf{A}$  represents the abundance of the endmember  $\mathbf{r}_i$  in pixel  $\tilde{\mathbf{s}}_j$ . The abundances obey the nonnegativity and sum-to-one constraints:  $\mathbf{A}_{ij} \geq 0$  for all  $i$  and  $j$ , and  $\sum_{i=1}^M \mathbf{A}_{ij} = 1$  for all  $j$ . Note that the tilde placed over symbols refers to noise-free data and all vectors are column vectors.

In this study, we shall assume that the endmembers are unknown but present in the scene. Let  $\omega$  be a subset of  $N'$  indexes in  $\{1, \dots, N\}$  that contains at least the column index of each endmember. Under these assumptions, and without loss of generality, we observe that the mixing model (1) can be reformulated as follows

$$\tilde{\mathbf{S}} = \tilde{\mathbf{S}}_\omega \mathbf{X} \quad (2)$$

where  $\tilde{\mathbf{S}}_\omega = (\tilde{\mathbf{s}}_{\omega_1}, \dots, \tilde{\mathbf{s}}_{\omega_{N'}})$  denotes the restriction of  $\tilde{\mathbf{S}}$  to its columns indexed by  $\omega$ , and  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_{N'})^\top$  is the abundance matrix. Similarly as above,  $\mathbf{X}_{ij}$  is the abundance of  $\tilde{\mathbf{s}}_{\omega_i}$  in  $\tilde{\mathbf{s}}_j$ . On the one hand, if  $\tilde{\mathbf{s}}_{\omega_i}$  is an endmember,  $\mathbf{x}_i$  has non-zero entries and represents the corresponding abundance map. On the other hand, if  $\tilde{\mathbf{s}}_{\omega_i}$  is a mixed pixel,  $\mathbf{x}_i$  has all its elements equal to zero. As a consequence,  $\mathbf{X}$  admits  $N' - M$  rows of zeros, the other rows being equal to rows of  $\mathbf{A}$ .

This means that  $\mathbf{X}$  allows to identify the endmembers in  $\tilde{\mathbf{S}}$  through its non-zero rows, which is an interesting property to be exploited in the case where the endmembers are unknown. Let us now turn to the more realistic situation where some noise corrupts the observations. In this case, model (2) becomes

$$\mathbf{S} = \tilde{\mathbf{S}} + \mathbf{E} = \tilde{\mathbf{S}}_\omega \mathbf{X} + \mathbf{E} \quad (3)$$

where  $\mathbf{S}$  denotes the available data and  $\mathbf{E}$  the noise.

The aim of this paper is to derive two unmixing approaches with increasing complexity, depending on how noise is to be handled. These methods are blind in the sense that the endmembers and their cardinality are unknown. The first one considers the approximate model

$$\mathbf{S} \approx \mathbf{S}_\omega \mathbf{X} + \mathbf{E} \quad (4)$$

Compared to (3), we thus assume that noise does not dramatically affect factorization of the mixing process, which is valid for very high signal-to-noise ratio (SNR). With this approach, we shall look for a few columns of  $\mathbf{S}_\omega$  that can effectively represent the whole scene. This strategy subserves a blind and self-dependent framework. It departs from methods based on a preselected dictionary of endmembers estimated from other experimental conditions, and thus do not accurately represent the endmembers in  $\mathbf{R}$ . In order to estimate the abundance matrix  $\mathbf{X}$ , we use prior information. First, we impose that the estimated abundances obey the non-negativity and sum-to-one constraints, namely,  $\mathbf{X}_{ij} \geq 0$  for all  $(i, j)$ , and  $\sum_{i=1}^N \mathbf{X}_{ij} = 1$  for all  $j$ . In addition, as discussed above, the algorithm has to force rows of  $\mathbf{X}$  to be zero vectors in order to identify the endmembers. Because the locations and the cardinality of the endmembers are unknown, the set of candidates has to be sufficiently large, that is,  $N' \gg M$ . We thus expect many rows in  $\mathbf{X}$  to be equal to zero. To promote this effect, the so-called Group Lasso  $\ell_{2,1}$ -norm regularization can be employed [9]. Because model (4) becomes a poor approximation of model (3) as the noise power increases, we shall also propose an alternative strategy to solve the unmixing problem based on the exact model (3). The first approach leads to a convex optimization problem that can be solved with the Alternating Direction Method of Multipliers (ADMM) [10]. The second one takes the noise in  $\mathbf{S}_\omega$  into account, which results in a non-convex and heteroscedastic optimization problem. The latter will be solved with an Iterative Reweighted Least Squares (IRLS) algorithm.

Few models of the form (3) have been studied in the literature [11]–[13]. These last three works assume that  $\mathbf{S}_\omega$  is noise-free. Moreover, in [11], the authors use an  $\ell_\infty$ -norm rather than the  $\ell_{2,1}$ -norm regularization considered here. In [12], the authors derive a Matching Pursuit approach [14] in order to estimate the endmembers. A similar technique is considered in [13], but the authors do not assume that the endmembers are present in the scene and use a predefined dictionary. Finally, note that the non-negativity and sum-to-one constraints are not considered in [11], [12].

The rest of this paper is organized as follows. Sections II and III respectively describe the unmixing models (4) and (3), and the corresponding estimation methods. Section IV provides experimental results on synthetic and real data. Finally, Section V concludes this paper.

## II. GROUP LASSO WITH UNIT SUM AND POSITIVITY CONSTRAINTS (GLUP)

### A. Model description

The aim of this section is to derive the estimation method for model (4), and finally define each step of the ADMM run to get the solution. In this approximate model, we assume that the noise  $\mathbf{E}$  is Gaussian independent and identically distributed, with zero mean and possibly unknown variance  $\sigma^2$ , that is,  $\mathbf{E}_{k,i} \sim \mathcal{N}(0, \sigma^2)$ . The negative log-likelihood for model (4) is given by

$$\mathcal{L}(\mathbf{X}) = \frac{NL}{2} \log(2\pi) + \frac{NL}{2} \log(\sigma^2) + \frac{1}{2\sigma^2} \|\mathbf{S} - \mathbf{S}_\omega \mathbf{X}\|_F^2 \quad (5)$$

The Maximum Likelihood (ML) estimate, namely, the minimizer of  $\mathcal{L}(\mathbf{X})$ , is the solution of the usual Least Squares (LS) approximation problem  $\min_{\mathbf{X}} \|\mathbf{S} - \mathbf{S}_\omega \mathbf{X}\|_F^2$ . Since model (4) follows from an approximation of model (3), the relevance of this LS fidelity term is essentially to ensure that  $\mathbf{S}_\omega \mathbf{X}$  matches  $\mathbf{S}$ . The unmixing problem under investigation, however, requires that  $\mathbf{X}$  only has a few rows different from zero, in addition to the non-negativity and sum-to-one constraints. This leads to following convex optimization problem

$$\begin{aligned} \min_{\mathbf{X}} \quad & \frac{1}{2} \|\mathbf{S} - \mathbf{S}_\omega \mathbf{X}\|_F^2 + \mu \sum_{k=1}^N \|\mathbf{x}_k\|_2 \\ \text{subject to} \quad & \mathbf{X}_{ij} \geq 0 \quad \forall i, j \\ & \sum_{i=1}^N \mathbf{X}_{ij} = 1 \quad \forall j \end{aligned} \quad (6)$$

with  $\mu \geq 0$  a regularization parameter and  $\mathbf{x}_k$  the  $k$ -th row of  $\mathbf{X}$ . The Group Lasso regularization term induces sparsity in the estimated abundance matrix at the group level [9], by possibly driving all the entries in several rows  $\mathbf{x}_k$  of  $\mathbf{X}$  to zero. It is worth noting that when  $\mu = 0$  and  $\mathbf{S}_\omega = \mathbf{S}$ , the solution of problem (6) is the identity matrix  $\mathbf{X} = \mathbf{I}$ . It follows that the efficiency of the approach relies on the  $\ell_{2,1}$ -norm regularization function.

### B. ADMM algorithm

The solution of problem (6) can be obtained in a simple and flexible manner using the ADMM algorithm [10]. We consider the canonical form

$$\begin{aligned} \min_{\mathbf{X}, \mathbf{Z}} \quad & \frac{1}{2} \|\mathbf{S} - \mathbf{S}_\omega \mathbf{X}\|_F^2 + \mu \sum_{k=1}^N \|\mathbf{z}_k\|_2 + \mathcal{I}(\mathbf{Z}) \\ \text{subject to} \quad & \mathbf{A}\mathbf{X} + \mathbf{B}\mathbf{Z} = \mathbf{C} \end{aligned} \quad (7)$$

with

$$\mathbf{A} = \begin{pmatrix} \mathbf{I} \\ \mathbf{1}^\top \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} -\mathbf{I} \\ \mathbf{0}^\top \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} \mathbf{0} \\ \mathbf{1}^\top \end{pmatrix},$$

where  $\mathcal{I}$  is the indicator of the positive orthant guarantying the positivity constraint, that is,  $\mathcal{I}(\mathbf{Z}) = 0$  if  $\mathbf{Z} \succeq \mathbf{0}$  and  $+\infty$  otherwise. The equality constraint imposes the consensus  $\mathbf{X} = \mathbf{Z}$  and the sum-to-one constraint. In matrix form, the augmented Lagrangian for problem (7) is given by [15]

$$\mathcal{L}_\rho(\mathbf{X}, \mathbf{Z}, \boldsymbol{\Lambda}) = \frac{1}{2} \|\mathbf{S} - \mathbf{S}_\omega \mathbf{X}\|_F^2 + \mu \sum_{k=1}^N \|\mathbf{z}_k\|_2 + \mathcal{I}(\mathbf{Z}) + \text{trace}(\boldsymbol{\Lambda}^\top (\mathbf{A}\mathbf{X} + \mathbf{B}\mathbf{Z} - \mathbf{C})) + \frac{\rho}{2} \|\mathbf{A}\mathbf{X} + \mathbf{B}\mathbf{Z} - \mathbf{C}\|_F^2$$

where  $\Lambda$  is the matrix of Lagrange multipliers,  $\mu$  and  $\rho$  are positive regularization and penalty parameters, respectively. The flexibility of the ADMM lies in the fact that it splits the initial variable  $\mathbf{X}$  into two variables,  $\mathbf{X}$  and  $\mathbf{Z}$ , and equivalently the initial problem into two subproblems. At iteration  $k + 1$ , the ADMM algorithm is outlined by three sequential steps:

1) *Minimization of  $\mathcal{L}_\rho(\mathbf{X}, \mathbf{Z}^k, \Lambda^k)$  with respect to  $\mathbf{X}$* : This step takes into account the previous estimates of  $\mathbf{Z}$  and  $\Lambda$ . The augmented Lagrangian is quadratic in terms of  $\mathbf{X}$ . As a result, the solution has an analytical expression that is obtained by setting the gradient of  $\mathcal{L}_\rho(\mathbf{X}, \mathbf{Z}^k, \Lambda^k)$  to zero:

$$\mathbf{X}^{k+1} = (\mathbf{S}_\omega^\top \mathbf{S}_\omega + \rho \mathbf{A}^\top \mathbf{A})^{-1} (\mathbf{S}_\omega^\top \mathbf{S} - \mathbf{A}^\top [\Lambda^k + \rho (\mathbf{B}\mathbf{Z}^k - \mathbf{C})]) \quad (8)$$

2) *Minimization of  $\mathcal{L}_\rho(\mathbf{X}^{k+1}, \mathbf{Z}, \Lambda^k)$  with respect to  $\mathbf{Z}$* : After removing the terms that are independent of  $\mathbf{Z}$ , the minimization of  $L_\rho(\mathbf{X}^{k+1}, \mathbf{Z}, \Lambda^k)$  with respect to  $\mathbf{Z}$  reduces to solving the following problem:

$$\begin{aligned} \min_{\mathbf{Z}} \quad & \mu \sum_{k=1}^N \|\mathbf{z}_k\|_2 + \text{trace}(\Lambda^\top \mathbf{B}\mathbf{Z}) + \frac{\rho}{2} \|\mathbf{A}\mathbf{X} + \mathbf{B}\mathbf{Z} - \mathbf{C}\|_F^2 \\ \text{subject to} \quad & \mathbf{Z} \succeq \mathbf{0} \end{aligned} \quad (9)$$

This minimization step can be split into  $N$  problems given the structure of matrices  $\mathbf{A}$  and  $\mathbf{B}$ , one for each row of  $\mathbf{Z}$ , that is,

$$\min_{\mathbf{z}} \quad \frac{1}{2} \|\mathbf{z} - \mathbf{v}\|_2^2 + \alpha \|\mathbf{z}\|_2 + \mathcal{I}(\mathbf{z}) \quad (10)$$

where  $\mathbf{v} = \mathbf{x} + \rho^{-1} \boldsymbol{\lambda}$ ,  $\alpha = \rho^{-1} \mu$ ,  $\boldsymbol{\lambda}$ ,  $\mathbf{x}$  and  $\mathbf{z}$  correspond to a row in  $\Lambda$ ,  $\mathbf{X}$  and  $\mathbf{Z}$  respectively. The minimization problem (10) admits a unique solution, given by the proximity operator [16] of  $f(\mathbf{z}) = \alpha \|\mathbf{z}\|_2 + \mathcal{I}(\mathbf{z})$

$$\begin{cases} \mathbf{z}^* = \mathbf{0} & \text{if } \|(\mathbf{v})_+\|_2 < \alpha \\ \mathbf{z}^* = \left(1 - \frac{\alpha}{\|(\mathbf{v})_+\|_2}\right) (\mathbf{v})_+ & \text{otherwise} \end{cases} \quad (11)$$

where  $(\cdot)_+ = \max(\mathbf{0}, \cdot)$ . On the one hand, the proximity operator of  $f_1(\mathbf{z}) = \alpha \|\mathbf{z}\|_2$  is the *Multidimensional Shrinkage Thresholding Operator* (MiSTO) [17]. On the other hand, the proximity operator of the indicator function  $f_2(\mathbf{z}) = \mathcal{I}(\mathbf{z})$  is the projection onto the positive orthant. The proximity operator of  $f(\mathbf{z})$  in (11), that we refer to as Positively constrained MiSTO, is an extension of both previous operators. The solution is of the form  $\text{prox}_f = \text{prox}_{f_1} \circ \text{prox}_{f_2}$ , that is, the thresholding of the projection. Operator (11) was recently used in [18]. A proof for this operator can be found in the Appendix.

3) *Update of the Lagrange multipliers  $\Lambda$* : Update of the Lagrange multipliers is carried out at the end of each iteration.  $\Lambda^{k+1}$  represents the running sum of residuals. It gives an insight on the convergence of the algorithm. As  $k$  tends to infinity, the primal residual tends to zero and  $\Lambda^{k+1}$  converges to the dual optimal point.

$$\Lambda^{k+1} = \Lambda^k + \rho(\mathbf{A}\mathbf{X}^{k+1} + \mathbf{B}\mathbf{Z}^{k+1} - \mathbf{C}). \quad (12)$$

As suggested in [10], a reasonable stopping criteria is that the primal and dual residuals must be smaller than some tolerance thresholds, namely,

$$\|\mathbf{A}\mathbf{X}^{k+1} + \mathbf{B}\mathbf{Z}^{k+1} - \mathbf{C}\|_2 \leq \epsilon_{\text{pri}} \quad \text{and} \quad \|\rho \mathbf{A}^\top \mathbf{B}(\mathbf{Z}^{k+1} - \mathbf{Z}^k)\|_2 \leq \epsilon_{\text{dual}} \quad (13)$$

The pseudocode for the so-called GLUP method is provided by Algorithm 1. It is worth emphasizing that the main difference between the ADMM steps developed in GLUP and those in [13] arises in the ADMM variable splitting. The global problem

in [13] is decomposed into three subproblems: the least squares minimization, the Group Lasso regularization, and projection on the positive orthant. A consequence is that three ADMM variables are used instead of two, which leads to additional steps. In addition, the sum-to-one constraint is not considered in [13].

---

**Algorithm 1** :  $X = \text{GLUP}(S, S_\omega, \rho, \mu)$ 


---

```

1: Precompute  $A$ ,  $B$ , and  $C$ 
2: Initialize  $Z = \mathbf{0}$  and  $\Lambda = \mathbf{0}$ 
3:  $Q = (S_\omega^\top S_\omega + \rho A^\top A)^{-1}$ 
4: while  $\|R\|_2 \geq \epsilon_{\text{pri}}$  or  $\|P\|_2 \geq \epsilon_{\text{dual}}$  do
5:    $X = Q(S_\omega^\top S - A^\top(\Lambda + \rho[BZ - C]))$ 
6:    $Z^{\text{old}} = Z$ 
7:   for  $i = 1 \dots N'$  do
8:      $v_i = ((x_i)^\top + \rho^{-1} \lambda_i)_+$ 
9:     if  $\|v_i\|_2 < \rho^{-1} \mu$  then
10:       $z_i = \mathbf{0}$ 
11:     else
12:       $z_i = \left(1 - \frac{\mu}{\rho \|v_i\|_2}\right) v_i$ 
13:     end if
14:   end for
15:    $R = AX + BZ - C$ 
16:    $P = \rho AB(Z - Z^{\text{old}})$ 
17:    $\Lambda = \Lambda + \rho(AX + BZ - C)$ 
18: end while

```

---

### III. REDUCED NOISE FOR GROUP LASSO WITH UNIT SUM AND POSITIVITY CONSTRAINTS (NGLUP)

#### A. Model description

We now turn to the more realistic model (3). Let  $E_\omega$  and  $I_\omega$  be the  $L$ -by- $N'$  and  $N$ -by- $N'$  restrictions of  $E$  and  $I$  to the columns indexed by  $\omega$ , respectively. The noisy mixing model (3) is given by

$$S = (S_\omega - E_\omega)X + E = S_\omega X + E(I - I_\omega X) \quad (14)$$

This model belongs to the family of heteroscedastic regression [19], where the variance of the additive noise depends on  $X$ .

Let us define the matrix  $C(X)$  as

$$C(X) = (I - I_\omega X)^\top (I - I_\omega X) \quad (15)$$

It follows that

$$\text{vec}(\mathbf{E}(\mathbf{I} - \mathbf{I}_\omega \mathbf{X})) \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{C}(\mathbf{X}) \otimes \mathbf{I}) \quad (16)$$

where  $\otimes$  represents the Kronecker product of matrices, and  $\text{vec}(\cdot)$  is the operator that stacks the columns of a matrix on top of each other. The presence of  $\mathbf{X}$  in the expression of the noise variance has consequences on the negative log-likelihood of model (14), which no longer leads to the LS approximation problem

$$\begin{aligned} \mathcal{L}(\mathbf{X}, \sigma^2) &= \frac{1}{2} \log |\sigma^2 \mathbf{C}(\mathbf{X}) \otimes \mathbf{I}| + \frac{1}{2} \text{vec}(\mathbf{S} - \mathbf{S}_\omega \mathbf{X})^\top (\sigma^2 \mathbf{C}(\mathbf{X}) \otimes \mathbf{I})^{-1} \text{vec}(\mathbf{S} - \mathbf{S}_\omega \mathbf{X}) \\ &= \frac{L}{2} \log |\sigma^2 \mathbf{C}(\mathbf{X})| + \frac{1}{2} \text{trace}((\mathbf{S} - \mathbf{S}_\omega \mathbf{X})(\sigma^2 \mathbf{C}(\mathbf{X}))^{-1}(\mathbf{S} - \mathbf{S}_\omega \mathbf{X})^\top) \\ &= \frac{L}{2} \log |\sigma^2 \mathbf{C}(\mathbf{X})| + \frac{1}{2} \|\mathbf{S} - \mathbf{S}_\omega \mathbf{X}\|_{(\sigma^2 \mathbf{C}(\mathbf{X}))^{-1}}^2 \end{aligned} \quad (17)$$

The ML estimate for problem (17) with the Group Lasso regularization, nonnegativity and sum-to-one constraints yields the following constrained optimization problem:

$$\begin{aligned} \min_{\mathbf{X}, \sigma^2} \quad & \frac{L}{2} \log |\sigma^2 \mathbf{C}(\mathbf{X})| + \frac{1}{2} \|\mathbf{S} - \mathbf{S}_\omega \mathbf{X}\|_{(\sigma^2 \mathbf{C}(\mathbf{X}))^{-1}}^2 + \mu \sum_{k=1}^N \|\mathbf{x}_k\|_2 \\ \text{subject to} \quad & \mathbf{X}_{ij} \geq 0 \quad \forall i, j \\ & \sum_{i=1}^N \mathbf{X}_{ij} = 1 \quad \forall j \end{aligned} \quad (18)$$

### B. Alternating ADMM algorithm

Problem (18) is not convex and requires the estimation of  $\sigma^2$ . The second term in the objective function is closely related to Iteratively Reweighted Least Squares (IRLS) algorithms used as a solution in heteroscedastic models [20]. Note that, in IRLS algorithms,  $(\mathbf{S} - \mathbf{S}_\omega \mathbf{X})$  in equation (17) is usually substituted by  $(\mathbf{S} - \mathbf{S}_\omega \mathbf{X})^\top$ . This has consequences on the  $\mathbf{X}$  minimization step. In IRLS, the estimation process is carried out in two steps. The first step consists of updating weights, which are usually set to be inversely proportional to variances. The second step is the calculation of the LS estimator using the updated weights. Many strategies can be used to estimate the variances for the weight matrix, see for example [19], [21], [22].

The resolution of problem (18) with respect to  $\sigma^2$  for fixed  $\mathbf{X}$  gives

$$\sigma^2(\mathbf{X}) = \frac{1}{NL} \text{trace}((\mathbf{S} - \mathbf{S}_\omega \mathbf{X}) \mathbf{C}(\mathbf{X})^{-1} (\mathbf{S} - \mathbf{S}_\omega \mathbf{X})^\top) \quad (19)$$

Let  $\mathbf{W}(\mathbf{X}) = \sigma^2(\mathbf{X}) \mathbf{C}(\mathbf{X})$  denote the weight matrix of the least squares term in (18). To solve problem (18) with respect to  $\sigma^2$  and  $\mathbf{X}$ , we propose to proceed iteratively. Let  $\mathbf{X}^k$  be the solution of the previous iteration. The first step consists of calculating  $\mathbf{W}(\mathbf{X}^k)$  using equations (15) and (19). In the second step, this updated weight matrix is used to estimate  $\mathbf{X}^{k+1}$  as follows

$$\begin{aligned} \min_{\mathbf{X}} \quad & \frac{1}{2} \|\mathbf{S} - \mathbf{S}_\omega \mathbf{X}\|_{(\mathbf{W}^k)^{-1}}^2 + \mu \sum_{k=1}^N \|\mathbf{x}_k\|_2 \\ \text{subject to} \quad & \mathbf{X}_{ij} \geq 0 \quad \forall i, j \\ & \sum_{i=1}^N \mathbf{X}_{ij} = 1 \quad \forall j \end{aligned} \quad (20)$$

where  $\mathbf{W}^k = \mathbf{W}(\mathbf{X}^k)$ . Given  $\mathbf{W}^k$ , problem (20) reduces to a weighted version of GLUP (6) due to the weighted norm in the first term. The ADMM solution developed in section II can be adapted to solve the optimization problem (20). Minimizing the

augmented Lagrangian with respect to  $\mathbf{Z}$ , and updating the Lagrange multipliers, can be carried out exactly as in Section II. For concision, only the  $\mathbf{X}$ -minimization step is described hereafter.

*Minimization of  $\mathcal{L}_\rho(\mathbf{X}, \mathbf{Z}^k, \mathbf{\Lambda}^k)$  with respect to  $\mathbf{X}$ :* Omitting the terms that do not depend on  $\mathbf{X}$ , the minimization of the augmented Lagrangian  $\mathcal{L}_\rho(\mathbf{X}, \mathbf{Z}^k, \mathbf{\Lambda}^k)$  with respect to  $\mathbf{X}$  leads to

$$\min_{\mathbf{X}} \quad \frac{1}{2} \|\mathbf{S} - \mathbf{S}_\omega \mathbf{X}\|_{(\mathbf{W}^k)^{-1}}^2 + \text{trace}(\mathbf{\Lambda}^\top (\mathbf{A}\mathbf{X})) + \frac{\rho}{2} \|\mathbf{A}\mathbf{X} + \mathbf{B}\mathbf{Z} - \mathbf{C}\|_F^2 \quad (21)$$

Problem (21) is quadratic in  $\mathbf{X}$  and admits an analytical solution obtained by setting the gradient to zero. This amounts to solving the Sylvester equation, which has an analytic solution [23]

$$\mathbf{S}_\omega^\top \mathbf{S}_\omega \mathbf{X} (\mathbf{W}^k)^{-1} + \rho \mathbf{A}^\top \mathbf{A} \mathbf{X} = \mathbf{S}_\omega^\top \mathbf{S} (\mathbf{W}^k)^{-1} - \rho \mathbf{A}^\top \left( \mathbf{B}\mathbf{Z}^k - \mathbf{C} + \frac{\mathbf{\Lambda}^k}{\rho} \right) \quad (22)$$

Problem (18) is not convex. An alternating optimization algorithm is more likely to converge to local minima with worse accuracy than the convex version. For this reason, we suggest, as a warm start, to initialize NGLUP with GLUP estimate. Algorithm 2 provides the pseudocode for NGLUP. The algorithm contains two main loops. The inner loop aims at finding the solution of problem (20), whereas the outer loop updates the least-square weight matrix.

#### IV. EXPERIMENTAL RESULTS

##### A. Synthetic Data

The performance of GLUP and NGLUP were evaluated using synthetic data. We used seven endmembers with 420 spectral samples extracted from the USGS library. Figure 1 shows the reflectance of the endmembers. The spectral mutual coherence between two spectra is defined as  $\theta_{ij} = \frac{\langle \mathbf{s}_i, \mathbf{s}_j \rangle}{\|\mathbf{s}_i\| \|\mathbf{s}_j\|}$ . The average mutual coherence of the eight endmembers was  $\theta_{\text{avg}} = 0.9171$ . The abundances were generated based on a Dirichlet distribution with unit parameter, as a consequence of which the resulting abundances obeyed the non-negativity and sum-to-one constraint, and were uniformly distributed over this simplex.

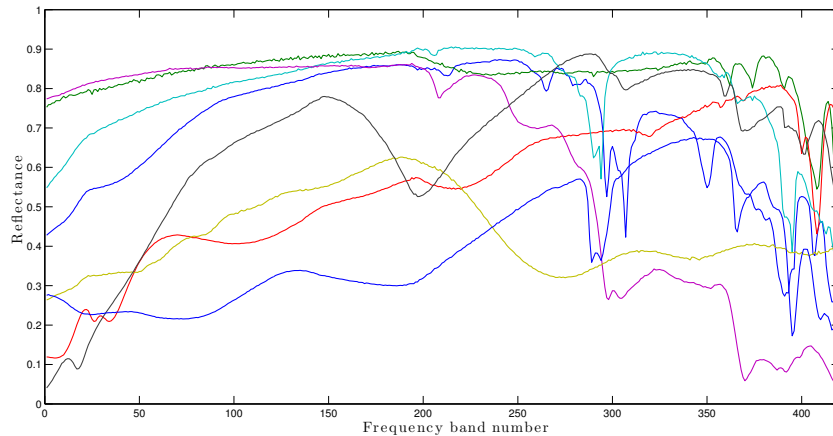


Fig. 1. Reflectance of selected endmembers from USGS Library.



---

**Algorithm 2** :  $\mathbf{X} = \text{NGLUP}(\mathbf{S}, \mathbf{S}_\omega, \rho^\circ, \mu^\circ, \rho, \mu)$ 


---

```

1: Precompute  $\mathbf{A}$ ,  $\mathbf{B}$ , and  $\mathbf{C}$ 
2: Initialize  $\mathbf{X} = \text{GLUP}(\mathbf{S}, \mathbf{S}_\omega, \rho^\circ, \mu^\circ)$ ,  $\mathbf{Z} = \mathbf{X}$ ,  $\mathbf{\Lambda} = \mathbf{0}$ 
3: while  $\|\mathbf{X} - \mathbf{X}_{\text{old}}\|_2 \geq \epsilon_{\text{tol}}$  do
4:    $\mathbf{C}(\mathbf{X}) = (\mathbf{I} - \mathbf{I}_\omega \mathbf{X})^\top (\mathbf{I} - \mathbf{I}_\omega \mathbf{X})$ 
5:    $\sigma^2(\mathbf{X}) = \frac{1}{NL} \text{trace}((\mathbf{S} - \mathbf{S}_\omega \mathbf{X}) \mathbf{C}(\mathbf{X})^{-1} (\mathbf{S} - \mathbf{S}_\omega \mathbf{X})^\top)$ 
6:    $\mathbf{W}(\mathbf{X}) = \sigma^2(\mathbf{X}) \mathbf{C}(\mathbf{X})$ 
7:    $\mathbf{X}^{\text{old}} = \mathbf{X}$ ,  $J = 1$ 
8:   while ( $\|\mathbf{R}\|_2 \geq \epsilon_{\text{pri}}$  or  $\|\mathbf{P}\|_2 \geq \epsilon_{\text{dual}}$ ) and ( $J \leq J_{\text{max}}$ ) do
9:      $\mathbf{X} = \text{solution of Sylvester equation (22)}$ 
10:     $\mathbf{Z}_{\text{old}} = \mathbf{Z}$ 
11:    for  $i = 1 \cdots N'$  do
12:       $\mathbf{v}_i = ((\mathbf{x}_i)^\top + \rho^{-1} \boldsymbol{\lambda}_i)_+$ 
13:      if  $\|\mathbf{v}_i\|_2 < \rho^{-1} \mu$  then
14:         $\mathbf{z}_i = \mathbf{0}$ 
15:      else
16:         $\mathbf{z}_i = \left(1 - \frac{\mu}{\rho \|\mathbf{v}_i\|_2}\right) \mathbf{v}_i$ 
17:      end if
18:    end for
19:     $\mathbf{R} = \mathbf{A}\mathbf{X} + \mathbf{B}\mathbf{Z} - \mathbf{C}$ 
20:     $\mathbf{P} = \rho \mathbf{A}\mathbf{B}(\mathbf{Z} - \mathbf{Z}_{\text{old}})$ 
21:     $\mathbf{\Lambda} = \mathbf{\Lambda} + \rho(\mathbf{A}\mathbf{X} + \mathbf{B}\mathbf{Z} - \mathbf{C})$ 
22:     $J = J + 1$ 
23:  end while
24: end while

```

---

First, we used three endmembers to generate an hyperspectral data set containing 100 pixels with a SNR of 50 dB. The pure pixels were indexed by integers 1–3 for simplicity, the mixed pixels being indexed by integers 4–100. We run GLUP algorithm using all the observations ( $\mathbf{S}_\omega = \mathbf{S}$ ) with  $\mu = 10$  and  $\rho = 100$ . The primal and dual tolerances were set to  $10^{-5}$ . Figure 2 shows the mean of each row  $\mathbf{x}_k$  of the estimated abundance matrix  $\hat{\mathbf{X}}$ . We observe that the first three pixels can be identified as the endmembers since the mean values of the first three rows are clearly different from zero. GLUP was able to provide this result in 4.73 seconds<sup>1</sup> with a Root Mean Square Error (RMSE), defined as  $\frac{1}{N^2} \|\hat{\mathbf{X}} - \mathbf{X}\|_{\text{F}}^2$ , equal to 0.0049.

<sup>1</sup>Machine specifications: 2.2 GHz Intel Core i7 processor and 8 GB RAM

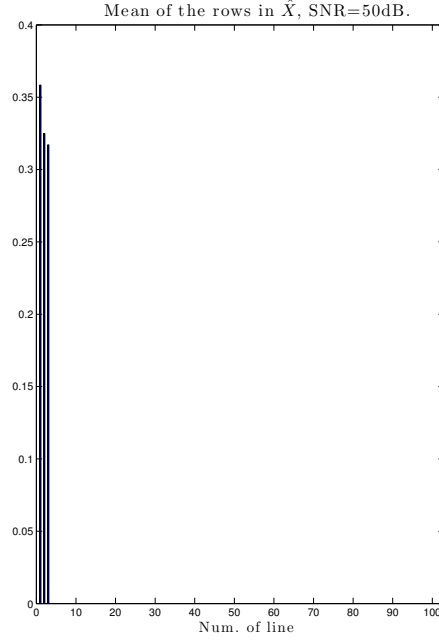


Fig. 2. Mean value of each row of  $\hat{\mathbf{X}}$  estimated with GLUP, obtained with 100 pixels and SNR = 50 dB.

TABLE I

PROBABILITY OF DETECTING  $\hat{M}$  ENDMEMBERS, USING SYNTHETIC DATA GENERATED WITH  $M = 7$  ENDMEMBERS.

| $M$               | 2           | 3    | 6    | 7           | 8    |
|-------------------|-------------|------|------|-------------|------|
| NGLUP<br>(30 dB)  | 0           | 0    | 0    | <b>0.98</b> | 0.02 |
| VD<br>(30 dB)     | <b>0.79</b> | 0.21 | 0    | 0           | 0    |
| NGLUP<br>( 20 dB) | 0           | 0    | 0.01 | <b>0.96</b> | 0.03 |
| VD<br>(20 dB)     | <b>1</b>    | 0    | 0    | 0           | 0    |

We tested NGLUP in less favorable conditions by increasing the number of endmembers and decreasing the SNR. To this end, 7 endmembers were used to generate 93 mixed pixels. Data were corrupted with an additive Gaussian noise with a SNR of 20 dB. We tested the algorithm for a maximum number of inner iterations  $J_{\max} = 1, 10$  and 100. We found that NGLUP converged to the same solution even when the number of inner iterations  $J$  was equal to 1. For this reason, only one inner iteration per outer iteration was used for the rest of the experiments. The running time of the algorithm was 45 seconds. Figure 3 shows the mean value of each row  $x_k$  of the abundance matrix  $\hat{\mathbf{X}}$  estimated by GLUP and NGLUP algorithms. In both cases, the 7 largest mean values correspond to the 7 endmembers. As expected, NGLUP converged to a sparser and more accurate solution than GLUP.

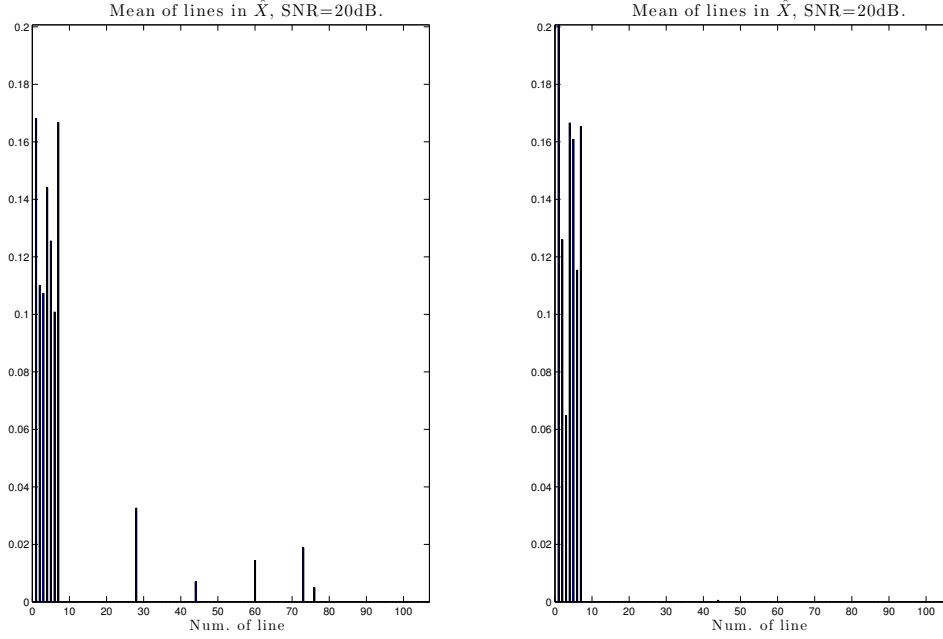


Fig. 3. Mean value of each row of  $\hat{\mathbf{X}}$ , obtained with 100 pixels and SNR = 20 dB: GLUP (left), NGLUP (right).

We repeated this simulation 100 times. For each realization, we examined the number of mean values of the rows of  $\hat{\mathbf{X}}$  that were larger than a predefined threshold equal to 0.01. We considered this value  $\hat{M}$  as the estimated number of endmembers in the scene. Table I provides the probability of detecting  $\hat{M}$  endmembers with our two approaches, given synthetic data generated with  $M = 7$  endmembers. The same task was performed using Virtual Dimensionality (VD) [2]. We compared the results of NGLUP with those of VD, the probability of false alarm of VD being set to  $10^{-3}$ . Table I shows that NGLUP was able to identify the presence of 7 endmembers in 98% (resp., 96%) of the cases with an SNR of 30 dB (resp. 20 dB). VD only identified 2 endmembers in most cases. Even with higher values of the SNR, VD did not identify the correct number of endmembers. This is due to the fact that VD has asymptotic convergence, and thus requires a very large number of observations in order to converge. This explains the poor performance of VD compared to NGLUP.

### B. Real data

In this section, we shall evaluate the performance of NGLUP using real hyperspectral data. The tests were performed on the so-called images of Pavia University,<sup>2</sup> provided by the ROSIS imaging spectrometer. The scene has a spatial dimension of  $610 \times 715$ , that is, a total of 207,400 pixels with a spatial resolution of 3.7 meters per pixel. Each pixel is composed of 102 spectral samples over the range 430-860 nm.

Given the high dimension of this data set, a subset  $\mathbf{S}$  of 300 pixels was randomly selected from the available observations. NGLUP was run with  $\mathbf{S}_w = \mathbf{S}$ . Based on this subset of observations, we selected those few that best described the whole scene.

<sup>2</sup>Available at <http://www.ehu.es/ccwintco/index.php/Home>

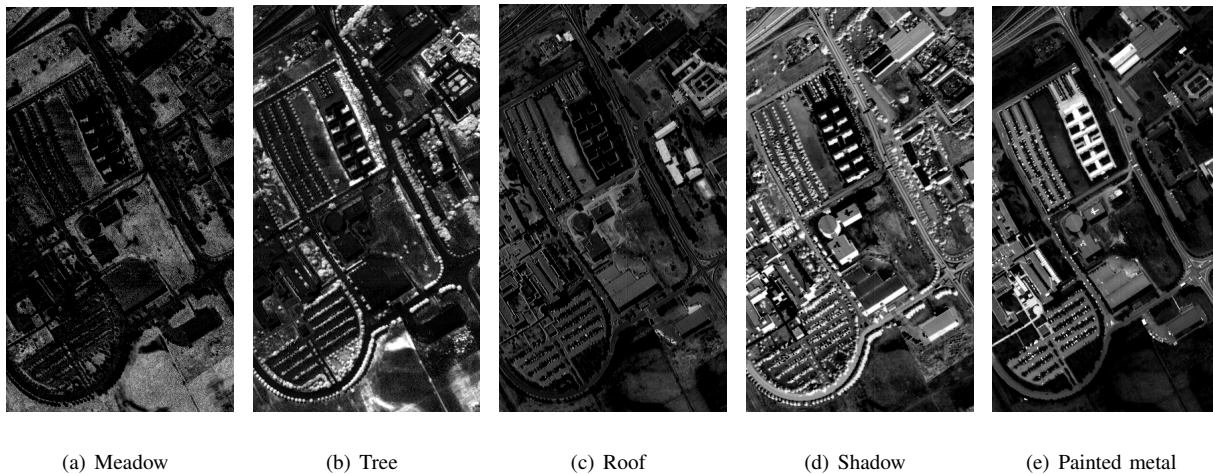


Fig. 4. Abundance maps of the five endmembers determined by NGLUP for Pavia University data set.

TABLE II  
PERFORMANCE OF NGLUP AND N-FINDR FOR PAVIA UNIVERSITY DATA SET

| Algorithm | RMSE   | max angle (rad) | avg angle (rad) |
|-----------|--------|-----------------|-----------------|
| NGLUP     | 0.0287 | 0.7468          | 0.074           |
| N-FINDR   | 0.0641 | 1.0592          | 0.1549          |

The estimated abundance matrix had a few rows different from zero, pointing out the candidate endmembers. Nevertheless, due to the extensive presence of redundant spectra, some rows revealed several occurrences of the same endmember. An additional step was thus required to remove redundant spectra among the endmembers determined by our algorithm. Several procedures have been proposed in the literature to perform this task. For example, in [11], the authors suggest to use K-means clustering in order to choose a subset of independent observations. In our experiments, we found it sufficient to impose a maximum value of 0.95 for the mutual coherence among the estimated endmembers. Redundant endmembers according to this criterion were discarded. With Pavia University data, this rule gave us 5 distinct endmembers.

We assumed that these endmembers, obtained from a small subset of the observations, were valid for the whole scene. This assumption can be justified by the fact that the image has lots of homogeneous surfaces where the spectral variability is negligible. We then used these endmembers and applied the Fully Constrained Least Squares (FCLS) algorithm on the whole data set. Figure 4 shows the abundance map for every endmember determined by NGLUP. The maps successfully describe the urban features of the scene and highlight its topography. They cast the pixels as combinations of meadow, tree, shadow, roof, painted metal sheet (with asphalt). Finally, we compared the performance of NGLUP with N-FINDR. Using the same subset of observations as NGLUP, we determined 5 endmembers with N-FINDR. Then, we applied FCLS on the whole image with the endmembers determined by N-FINDR. Table II shows the RMSE, the maximum and average spectral angles obtained for both methods. This comparison shows that NGLUP outperformed N-FINDR.

## V. CONCLUSION AND PERSPECTIVES

In this work, we presented two approaches for blind and fully constrained unmixing. Both methods are based on mixing models with increasing complexity, and allow to simultaneously determine the endmembers and estimate their local abundance in the scene. Compared to the first model called GLUP, the second model NGLUP explicitly considers that endmembers present in the scene are corrupted by noise. Experiments on synthetic and real data demonstrated the excellent performance of both approaches. Future work includes their extension to an online framework, which would allow to reduce their complexity and to make them adaptive to changing environmental conditions.

## APPENDIX

*Proof:* Since problem (11) is convex, we simply have to check the validity of the solution in the two cases  $\|(\mathbf{v})_+\|_2 > \alpha$  and  $\|(\mathbf{v})_+\|_2 < \alpha$ . Let  $f_0(\mathbf{z}) = \frac{1}{2}\|\mathbf{z} - \mathbf{v}\|_2^2 + \alpha\|\mathbf{z}\|_2$ . For  $\|(\mathbf{v})_+\|_2 > \alpha$ , the gradient of  $f_0$  is given by

$$\nabla f_0(\mathbf{z}^*) = \left(1 + \frac{\alpha}{\|\mathbf{z}^*\|_2}\right) \mathbf{z}^* - \mathbf{v} \quad (23)$$

Replacing by the appropriate expression from (11) yields

$$\nabla f_0(\mathbf{z}^*) = (\mathbf{v})_+ - \mathbf{v} \geq 0 \quad (24)$$

$$\mathbf{z}_i^* \cdot \nabla f_0(\mathbf{z}^*)_i \propto ((\mathbf{v})_+)_i \cdot ((\mathbf{v})_+ - \mathbf{v})_i = 0 \quad (25)$$

These two conditions correspond the optimality conditions, which means that  $\mathbf{z} \succeq \mathbf{0}$  is a solution for the constrained problem. For more details, refer to section 4.2.3 in [24].

For the second case, note that for every  $\mathbf{z} \succeq \mathbf{0}$ , we have

$$\sum_i \mathbf{z}_i \mathbf{v}_i \leq \sum_i \mathbf{z}_i (\mathbf{v}_i)_+ \leq \|\mathbf{z}\|_2 \cdot \|(\mathbf{v})_+\|_2 \quad (26)$$

It follows that

$$\begin{aligned} f_0(\mathbf{z}) - f_0(\mathbf{0}) &= \frac{1}{2} \sum_i \mathbf{z}_i^2 - \sum_i \mathbf{z}_i \mathbf{v}_i + \alpha\|\mathbf{z}\|_2 \\ &\geq \frac{1}{2} \|\mathbf{z}\|_2^2 - \|\mathbf{z}\|_2 \cdot \|(\mathbf{v})_+\|_2 + \alpha\|\mathbf{z}\|_2 \\ &\geq \frac{1}{2} \|\mathbf{z}\|_2^2 + \|\mathbf{z}\|_2 (\alpha - \|(\mathbf{v})_+\|_2) \end{aligned} \quad (27)$$

This proves that for  $\|(\mathbf{v})_+\|_2 \leq \alpha$ , the minimum is reached for  $\mathbf{z}^* = \mathbf{0}$ . ■

## REFERENCES

- [1] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 44–57, 2002.
- [2] C. Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 3, pp. 608–619, March 2004.
- [3] M. E. Winter, "N-FINDR: an algorithm for fast autonomous spectral endmember determination in hyperspectral data," in *Proc. SPIE Imaging Spectrometry*, October 1999.

- [4] D. C. Heinz and C. I. Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 3, pp. 529–545, March 2001.
- [5] O. Eches, N. Dobigeon, C. Mailhes, and J.-Y. Tourneret, "Bayesian estimation of linear mixtures using the normal compositional model," *IEEE Transaction on Image Processing*, vol. 19, no. 6, pp. 1403–1413, June 2010.
- [6] P. Honeine and C. Richard, "Geometric unmixing of large hyperspectral images: A barycentric coordinate approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 6, pp. 2185–2195, June 2012.
- [7] L. Miao, H. Qi, and H. Szu, "A maximum entropy approach to unsupervised mixed-pixel decomposition," *IEEE Transaction on Image Processing*, vol. 16, no. 4, pp. 1008–1021, April 2007.
- [8] Z. yang, G. Zhou, S. Xie, S. Ding, J.-M. Yang, and J. Zhang, "Blind spectral unmixing based on sparse nonnegative matrix factorization," *IEEE Transaction on Image Processing*, vol. 20, no. 4, pp. 1112–1125, April 2011.
- [9] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B (Statistical methodology)*, vol. 68, no. 1, pp. 49–67, February 2006.
- [10] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [11] E. Esser, M. Moller, S. Osher, G. Sapiro, and J. Xin, "A convex model for nonnegative matrix factorization and dimensionality reduction on physical space," *IEEE Transaction on Image Processing*, vol. 21, no. 7, pp. 3239–3252, July 2012.
- [12] X. Fu, W. Ma, T. Chan, J. Bioucas-Dias, and M. Iordache, "Greedy algorithms for pure pixels identification in hyperspectral unmixing: A multiple-measurement vector viewpoint," in *Proc. EUSIPCO*, 2013.
- [13] D. Iordache, J. Bioucas-Dias, and A. Plaza, "Collaborative sparse regression for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 1, pp. 341–354, February 2013.
- [14] Z. Zhang S. G. Mallat, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, december 1993.
- [15] J. Eckstein and D. P. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Mathematical Programming*, vol. 55, no. 1, pp. 293–318, 1992.
- [16] P. L. Combettes and J. C. Pesquet, "Proximal splitting methods in signal processing," in *arXiv:0912.3522*, 2009.
- [17] A. T. Puig, A. Wiesel, G. Fleury, and A. O. Hero, "Multidimensional shrinkage-thresholding operator and group lasso penalties," *IEEE Signal Processing Letters*, vol. 18, no. 6, pp. 363–366, April 2011.
- [18] E. Thiebaut, F. Soulez, and L. Denis, "Exploiting spatial sparsity for multiwavelength imaging in optical interferometry," *Journal of the Optical Society of America A*, vol. 30, no. 2, pp. 160–170, February 2013.
- [19] P. M. Hooper, "Iterative weighted least squares estimation in heteroscedastic linear models," *Journal of the American Statistical Association*, vol. 88, no. 421, pp. 179–184, 1993.
- [20] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Gntkr, "Iteratively reweighted least squares minimization for sparse recovery," *Communications on Pure and Applied Mathematics*, vol. 63, no. 1, pp. 1–38, 2010.
- [21] W. A. Fuller and J. N. K. Rao, "Estimation for a linear regression model with unknown diagonal covariance matrix," *The Annals of Statistics*, vol. 6, no. 5, pp. 1149–1158, september 1978.
- [22] R. J. Carroll and D. B. H. Cline, "An asymptotic theory for weighted least-squares with weights estimated by replication," *Biometrika*, vol. 75, no. 1, pp. 35–43, March 1988.
- [23] R. H. Bartels and G. W. Stewart, "Solution of the matrix equation  $AX + XB = C$ ," *Communications of the ACM*, vol. 15, no. 9, pp. 820–826, 1972.
- [24] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2008.